

新零售无人智能售货机商务数据分析

目录

一、	背景介绍.....	3
1、	行业介绍.....	3
2、	分析目标.....	3
二、	数据说明.....	3
1、	数据来源.....	3
2、	数据量.....	3
3、	数据格式.....	4
三、	数据处理.....	4
1、	检测并去除重复值.....	4
2、	检测并去除缺失值.....	4
3、	检测并去除异常值.....	4
四、	商品数据提取与订单量交易额分析.....	4
1、	提取数据（任务 1.1）.....	5
2、	交易额和订单量分析（任务 1.2）.....	5
3、	每单平均交易额及日均订单量分析（任务 1.3）.....	5
五、	商品数据可视化分析.....	8
1、	柱状图分析（任务 2.1）.....	8
2、	折线图及增长率分析（任务 2.2）.....	8
3、	饼图分析（任务 2.3）.....	10
4、	气泡图分析（任务 2.4）.....	10
5、	热力图分析（任务 2.5）.....	11
六、	自动售货机画像分析.....	13
1、	生成标签（任务 3.1）.....	13
2、	生成画像（任务 3.2）.....	14
七、	业务预测.....	16
1、	对 2018 年 1 月份交易额进行预测.....	16
2、	预测原理及结果可行性分析（任务 4.1）.....	16
3、	数据要求（任务 4.2）.....	17
八、	小结.....	18

一、 背景介绍

1、 行业介绍

自动售货机以线上经营的理念，提供线下的便利服务，以小巧、自助的经营模式节省人工成本，让实惠、高品质的商品触手可及，成为当下零售经营的又一主流模式。自动售货机内商品的供给频率、种类选择、供给量、站点选择等是自动售货机运营者需要重点关注的问题。因此，科学的商业数据分析能够帮助经营者了解用户需求，掌握商品需求量，为用户提供精准贴心的服务，是掌握经营方向的重要手段，对自动售货机这一营销模式的发展有着非常重要的意义

某商场在不同地点安放了 5 台自动售货机，编号分别为 A、B、C、D、E。

2、 分析目标

（1）根据自动售货机的经营特点，对经营指标数据、商品销售数据及市场需求进行分析，完成对销量、库存、盈利三个方面各项指标的计算，按要求绘制对应图表。

（2）为每台售货机所销售的商品贴上标签，使其能够很好地展现销售商品的特征。

（3）判断是否能够预测每台售货机的销售额，能的话给出结果，无法则给出理由。

二、 数据说明

1、 数据来源

数据来源：泰迪云平台提供

数据时间：2017 年 1 月 1 日至 2017 年 12 月 31 日

数据情况：附件 1 提供了每台自动售货机的商品销售数据

附件 2 提供了商品的分类

2、 数据量

（1）附件 1：订单量：70680 个

（2）附件 2：商品种类：315 件

3、 数据格式

(1) 附件 1:

订单号	设备 ID	应付金额	实际金额	商品	支付时间	地点	状态
DD201708167493 663618499909784	E43A6E078A07631	4.5	4.5	68g 好丽友 巧克力派 2 枚	2017/1 /10:53	D	已出货未退款

表 1 附件 1 格式

(2) 附件 2:

商品	大类	二级类
100g*5 瓶益力多	饮料	乳制品

表 2 附件 2 格式

三、 数据处理

1、 检测并去除重复值

由于给予的数据是每台售货机的每个商品的销售信息，因此对于重复数据，不需要去除的操作，因此检测并去除重复值完成。

2、 检测并去除缺失值

通过 Python，利用 isnull 或 notnull 方法对数据进行检测并未发现缺失值，操作完成

3、 检测并去除异常值

通过 Python，利用 3σ 原则或箱线图对数据进行检测未发现存在异常值，但在进行日期转换处理的时候出现错误，因此对数据进行再次检测，发现存在与众不同的时间表达方式，如表 3 因此对该数据，使用删除法去除。

订单号	设备 ID	应付金额	实际金额	商品	支付时间	地点	状态
DD2017081674932 41554692026752	E43A6E078A04228	4	4	55g 奥利奥 原味芝士饼 干	2017/2/29 3:44:00 PM	C	已出货未退款

表 3 异常值展示

四、商品数据提取与订单量交易额分析

1、提取数据（任务 1.1）

根据任务 1.1，按照附件 1 的数据，提取每台售货机对应的销售数据，放入 csv，分别命名。

2、交易额和订单量分析（任务 1.2）

根据任务 1.2，计算出每台售货机 2017 年各个月份的交易额、订单量，及所有售货机交易总额和订单总量。报告以 2017 年 5 月份为例，如表 4 所示。

售货机名称	订单量/个	交易额/元
A	564	2392.1
B	869	3681.2
C	756	3385.1
D	789	3729.4
E	1292	5699
总数	70680	286983.7

表 4 每台售货机交易额和订单量

从表格可以看出，售货机 E 的订单量和交易额都是明显大于其他的售货机的，说明售货机 E 的销量是比较好的，而售货机 A 的销量是五台售货机之中最差的。

因此可以加大对售货机 E 的投资和关注，增加销售额。

3、每单平均交易额及日均订单量分析（任务 1.3）

根据任务 1.3，计算出每台售货机每月的每单平均交易额与日均单量。如表 5、表 6、表 7、表 8、表 9 所示。

月份	日均单量	每单平均交易额
一月	354	4.680226
二月	258	3.638372
三月	350	4.305714
四月	895	4.159888
五月	1292	4.410991
六月	2593	3.817856

七月	813	3.919311
八月	1767	3.804471
九月	4134	4.125375
十月	2777	3.676125
十一月	5020	4.283227
十二月	3252	4.168973

表 5 A 售货机每单平均交易额和日均订单量

月份	日均单量	每单平均交易额
一月	366	3.753005
二月	185	3.255676
三月	265	3.614717
四月	603	4.07529
五月	869	4.236133
六月	1856	4.06805
七月	345	4.401739
八月	981	3.5842
九月	1745	4.130258
十月	2026	4.11234
十一月	2031	4.268784
十二月	2210	3.667014

表 6 B 售货机每单平均交易额和日均订单量

月份	日均单量	每单平均交易额
一月	335	4.506567
二月	114	3.864035
三月	255	3.58549
四月	447	4.036913
五月	756	4.477646
六月	1669	4.047394
七月	476	4.097689
八月	666	3.358709
九月	1040	4.307212
十月	1565	4.020703
十一月	1160	4.471552
十二月	2003	3.787868

表 7 C 售货机每单平均交易额和日均订单量

月份	日均单量	每单平均交易额
一月	379	4.328496
二月	207	3.826087
三月	263	3.769962
四月	734	4.403678
五月	789	4.726743
六月	1882	4.5017
七月	764	3.988351
八月	1259	3.913582
九月	1678	4.427294
十月	2216	4.27333
十一月	1943	4.352393
十二月	2379	3.943043

表 8 D 售货机每单平均交易额和日均订单量

月份	日均单量	每单平均交易额
一月	354	4.680226
二月	258	3.638372
三月	350	4.305714
四月	895	4.159888
五月	1292	4.410991
六月	2593	3.817856
七月	813	3.919311
八月	1767	3.804471
九月	4134	4.125375
十月	2777	3.676125
十一月	5020	4.283227
十二月	3252	4.168973

表 9 E 售货机每单平均交易额和日均订单量

从五张表格的数据可以看出，五台售货机的订单量在五六月份，以及九月到十二月份的订单量较其他的月份明显增加，由此可以得出，售货机商品在这个五个月份比较热销。

原因可能为五六月份天气较为炎热，居民购买饮料补充水分。而进入秋冬季之后，居民不愿前往较远处采购商品可能是秋冬季销量较多的原因。

五、 商品数据可视化分析

1、 柱状图分析（任务 2.1）

根据任务 2.1 绘制 2017 年每个月份销量前五商品柱状图，以 2017 年 6 月份为例。如图 1 所示。

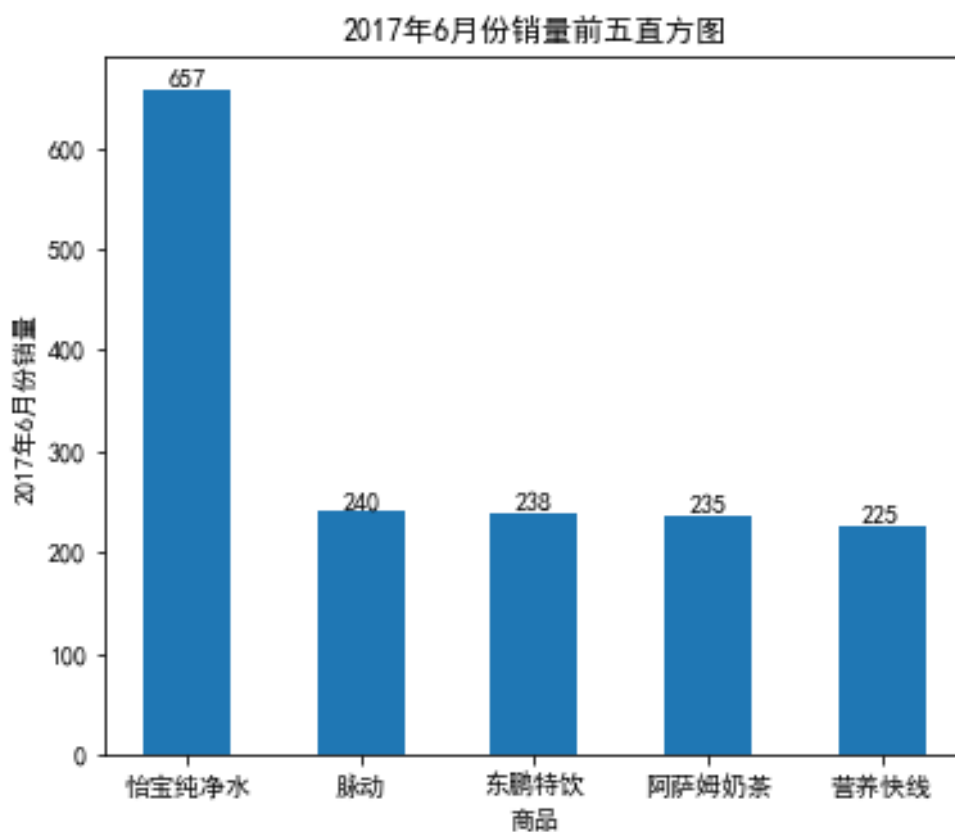


图 1 2017 年 6 月份销量前五直方图

观察图片可以发现，在售货机销量前五的商品基本都是属于饮料类商品，其中怡宝纯净水的数量明显领先。

由此也可以发现，自动售货机对于居民来说主要是平时购买饮料所需，又因为纯净水较多，可以猜测居民使用自动售货机的原因多为补充水分的原因。

2、 折线图及增长率分析（任务 2.2）

绘制每台售货机每月总交易额折线图及交易额月环比增长率柱状图，如图 2、图 3。

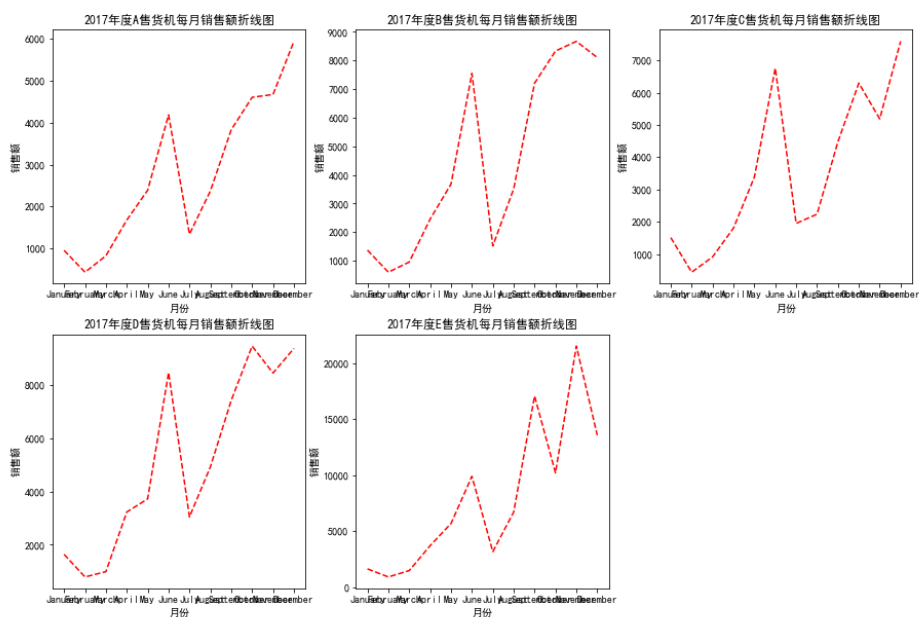


图 2 2017 年各台售货机每月销售额折线图

从图 2 也可发现，五台售货机在五六月份左右销量到达一个高峰，之后突然滑落，又在进入冬季之后逐渐上升，直到十一月十二月达到销量的顶点。

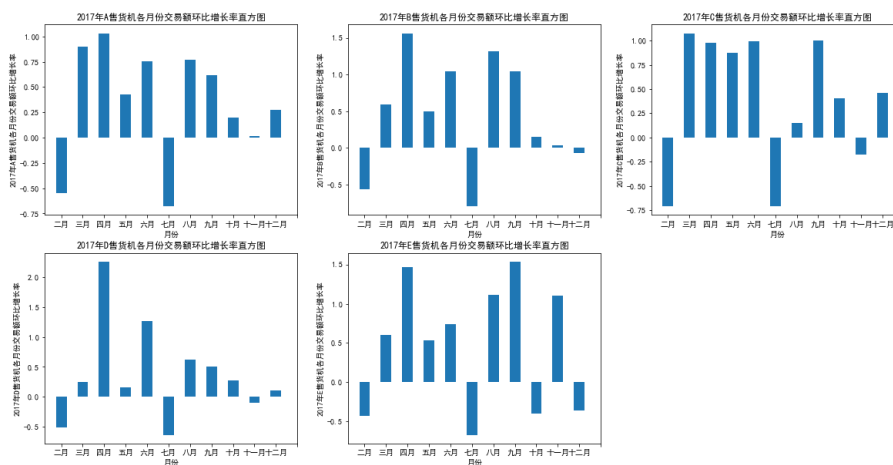


图 3 2017 年各台售货机各月份交易额环比增长率直方图

从图 3 可以发现，五台售货机在一月份和七月份的环比交易额增长率都是负数，说明交易额出现了负增长，由此也可看出六月份和十二月份的销量达到了高峰。

而单就增长率而言，二月份和九月份的增长率非常高，说明上半年销量在二月份开始逐渐增长直到六月份到达上半年的顶点，而下半年的销量从在九月份之后暴增之后，增加的 not 太多，维持了一个相对平稳的增长水平。

3、 饼图分析（任务 2.3）

绘制每台售货机毛利润占总毛利润比例的饼图（假设饮料类毛利润为 25%，非饮料类毛利润为 20%），如图 4。

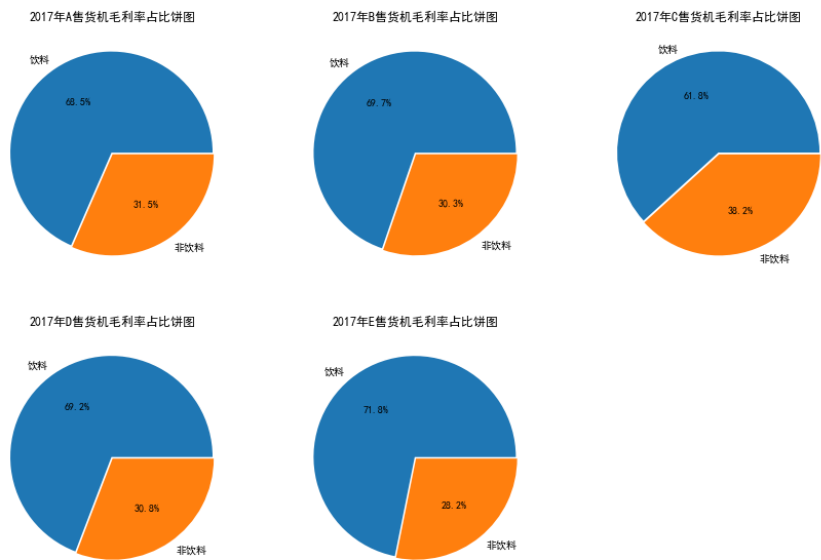


图 4 2017 年各台售货机毛利润占比饼图

由图 4 可以看出五载售货机中都是饮料类粘毛利润的比例较大，都在 60%–70%之间，饮料类的毛利润高。之前的图像显示，饮料类商品销量多，本图显示毛利润也多，因此售货机应该多贩卖饮料类商品。

4、 气泡图分析（任务 2.4）

绘制每月交易额均值气泡图。如图 5、图 6、图 8、图 7。

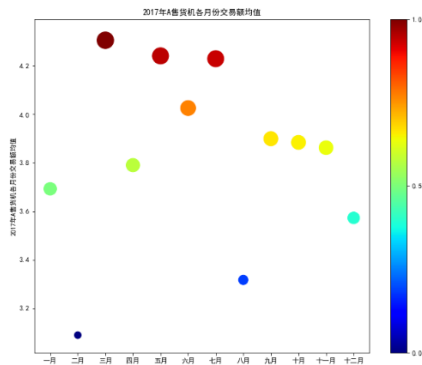


图 5 2017 年 A 售货机各月份交易额均值气泡图

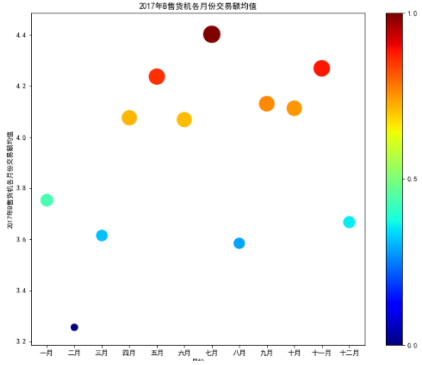


图 6 2017 年 B 售货机各月份交易额均值气泡图

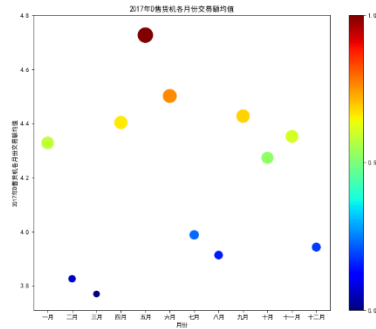


图 8 2017 年 C 售货机各月份交易额均值气泡图

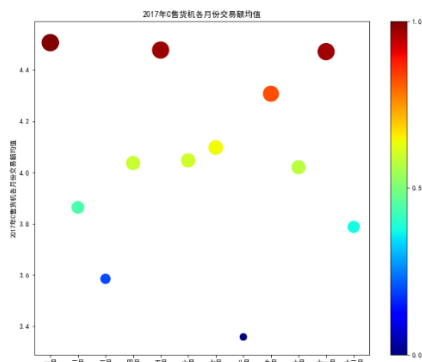


图 7 2017 年 D 售货机各月份交易额均值气泡图

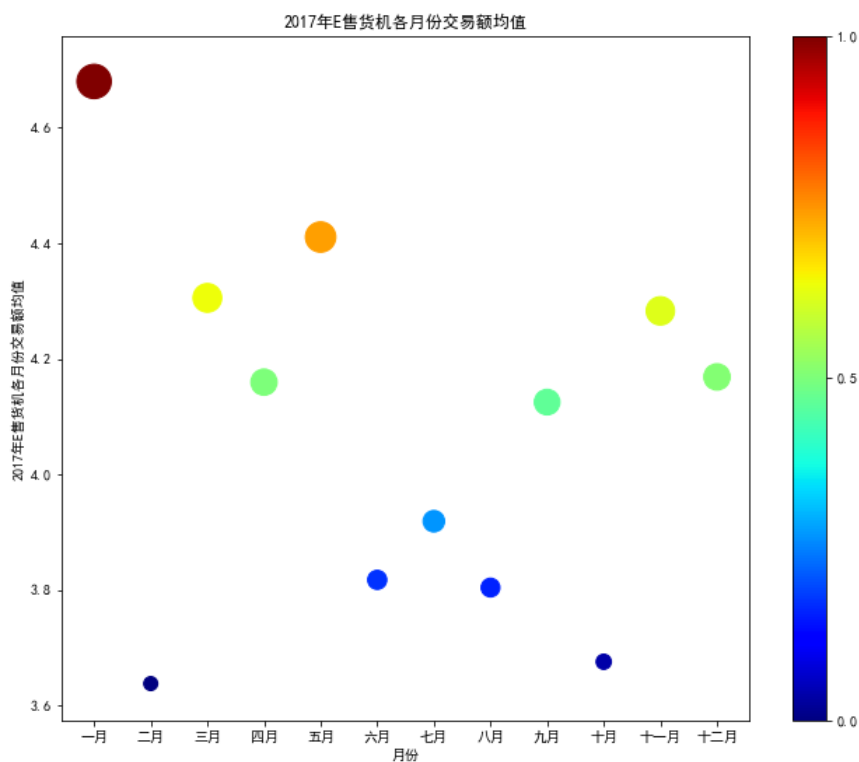


图 9 2017 年 E 售货机各月份交易额均值气泡图

图片展示了五台售货机分别的每月交易额均值气泡图。

5、热力图分析（任务 2.5）

绘制售货机 C 6、7、8 月份订单量热力图，如图 10、图 11、图 12。

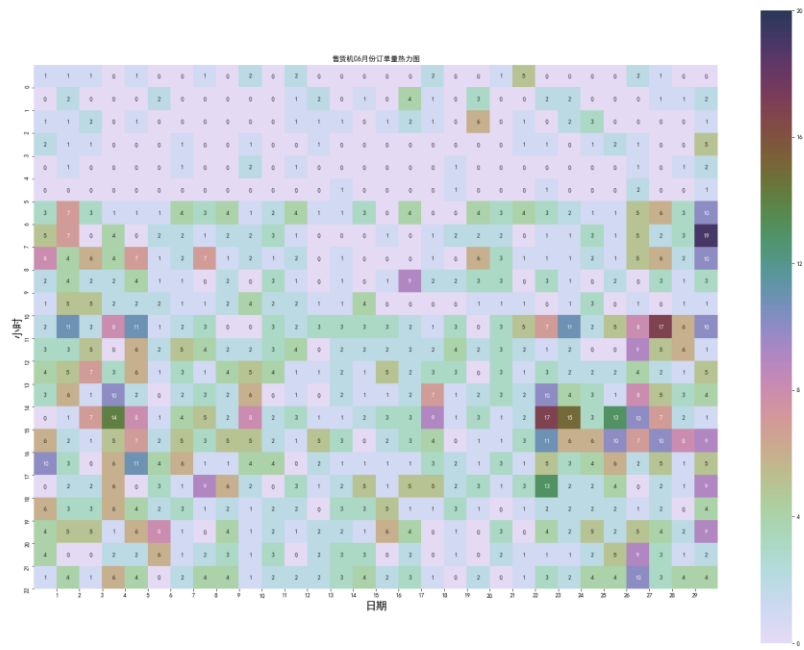


图 10 售货机 C 6 月份订单量热力图

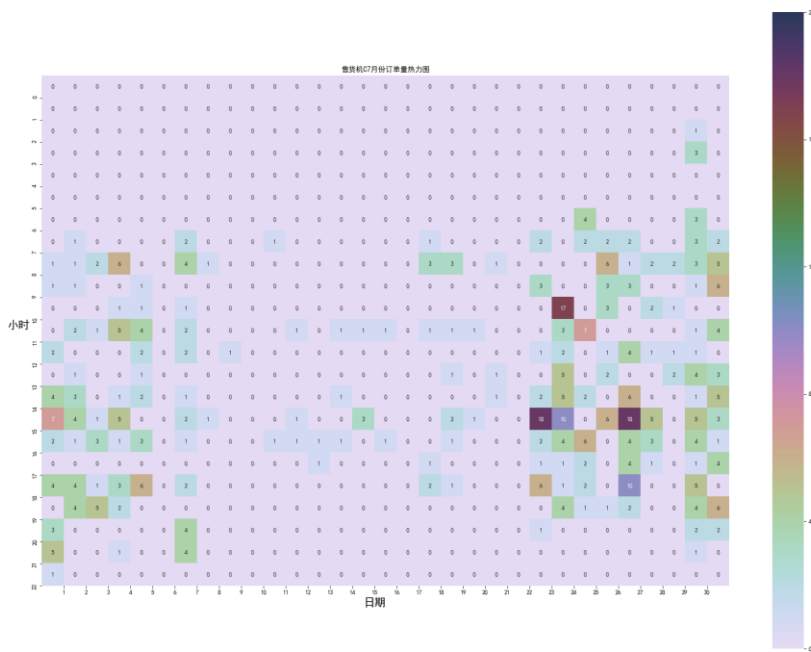


图 11 售货机 C 7 月份订单量热力图

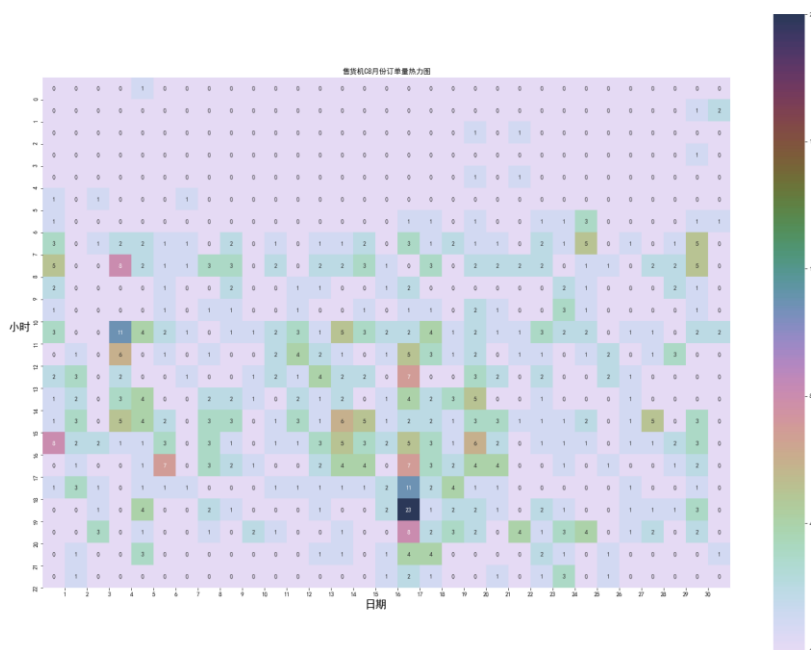


图 12 售货机 C 8 月份订单量热力图

热力图的横坐标为日期，纵坐标为小时，通过两者交集的不同颜色的深度，可以发现，6,7 月份销量较多的时间集中于月末，20 号之后。这两个月每天的销售高峰期则集中于中午时间段，而在 8 月份销量多集中于月中的时候，每日销量的巅峰也在傍晚时间。

从 6，7 月份的数据可以分析售货机销量较多的时间段为每天最为炎热的时间段，因此许多居民会选择购买饮料补充水分，这与前文的图相互映衬。也有可能是因为中午的时间靠近饭点，许多居民选择自动售货机的商品作为午餐选择，方便省时，而 8 月份的傍晚的高销量也可如此解释。

六、 自动售货机画像分析

1、 生成标签（任务 3.1）

给出各台销售机饮料类商品的标签，放入 csv，分别命名。

标签规则：

m=各月份销量均值

销量>m	热销
$0.5*m < \text{销量} < m$	正常
销量<0.5*m	滞销

表 10 标签规则表

2、生成画像（任务 3.2）

(1) 生成各台售货机画像，放入 csv，分别命名。

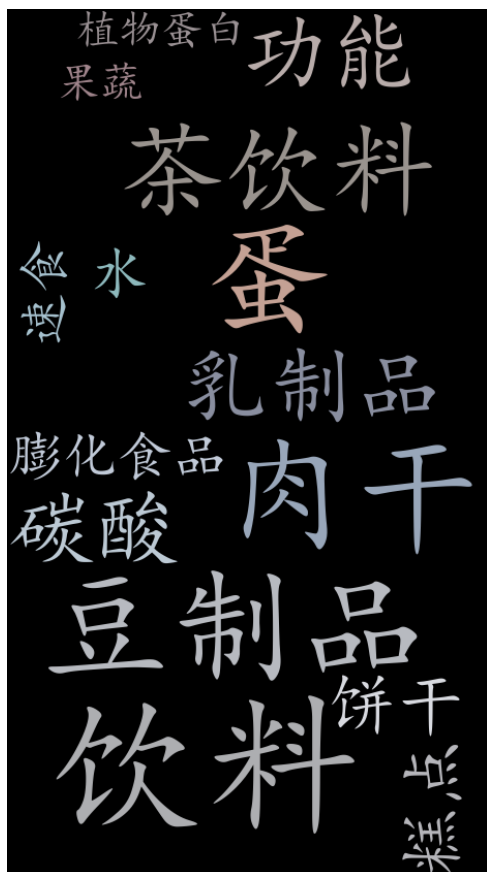


图 13 A 售货机画像

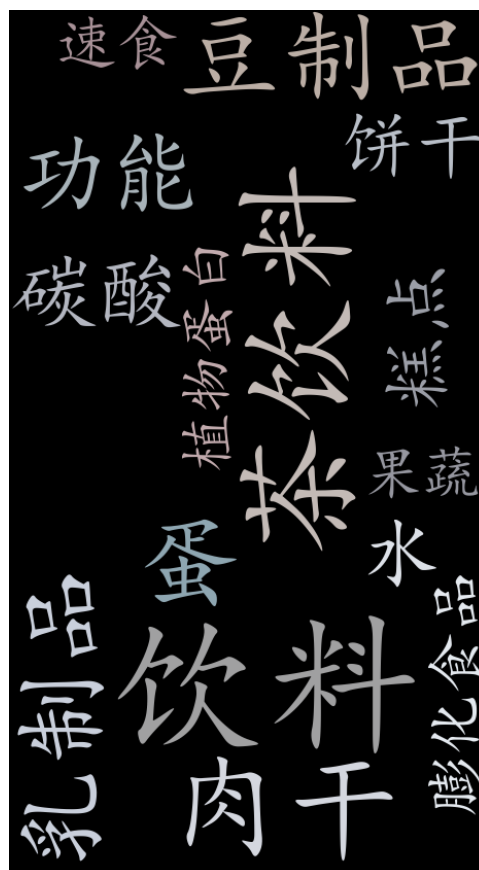


图 14 B 售货机画像



图 15 C 售货机画像

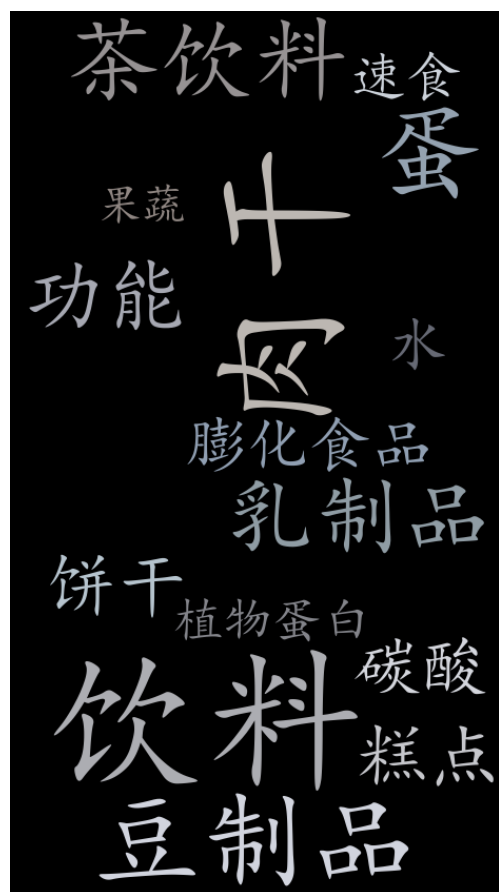


图 16 D 售货机画像

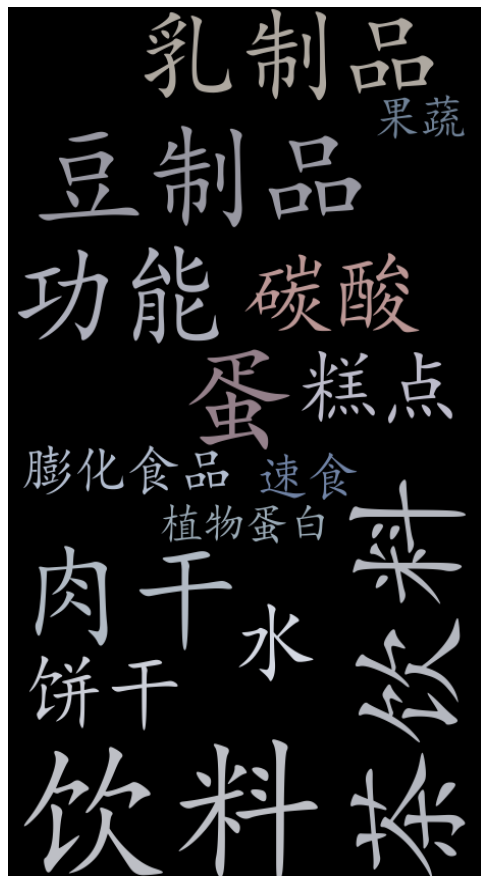


图 17 E 售货机画像

以上画像是把自动售货机的形状作为底图展示画像的。标签设置为商品二级类，各台售货机的标签组成画像，字体越大，则在该台售货机中，该类商品销量越多应增加采购放入自动售货机。

（2）营销意见

售货机 A：应采购更多的茶饮料，碳酸饮料，蛋制品，豆制品，和肉干类食品。

售货机 B：应采购更多的茶饮料，碳酸饮料，豆制品，乳制品，和肉干类食品。

售货机 C：应采购更多的肉干，豆制品，蛋制品，茶饮料，碳酸饮料，和乳制品。

售货机 D：应采购更多的茶饮料，碳酸饮料，蛋制品，豆制品，乳制品，水和肉干类食品。

售货机 E：应采购更多的茶饮料，碳酸饮料，蛋制品，豆制品，乳制品，和肉干类食品。

七、 业务预测

1、 对 2018 年 1 月份交易额进行预测

预测方法：由于只存在时间以及销量数据，因此采取时间序列预测的方法，采用 ARIMA 方法进行预测。

预测结果：表 11

2018 年 1 月交易额	饮料类	非饮料类
售货机 A	1030	399
售货机 B	1638	433
售货机 C	1071	363
售货机 D	1736	519
售货机 E	3009	601

表 11 2018 年 1 月份交易额预测

2、 预测原理及结果可行性分析（任务 4.1）

（1）预测原理：时间序列预测法其实是一种回归预测方法，属于定量预测，其基本原理是：一方面承认事物发展的延续性，运用过去的时间序列数据进行统计分析，推测出事物的发展趋势；另一方面充分考虑到由于偶然因素影响而产生的随机性，为了消除随机波动产生的影响，利用历史数据进行统计分析，并对数据进行适当处理，进行趋势预测。

（2）可行性分析：

①由于数据数量的限制，若采取每月预测，则数据数量不足，预测结果失真，因此采取每日预测。使用每日预测之后，因为需要预测未来 1 个月的数据，由于时间序列预测模型本

身的限制，在进行长期预测之时，越长时间的预测失真越严重，因此数据存在不可靠性。

②在进行 ARIMA 方法进行预测之时，无法寻找到完美的 p 值和 q 值，构建的 ARIMA 法中存在 p 值大于 0.05 的情况，因此备择假设，结果失真。

③未获得 2018 年 1 月份的交易数据对模型进行训练和预测，无法对模型的可靠性进行评价，模型的可用度不高。

3、 数据要求（任务 4.2）

- （1）增加数据量，确保模型的准确度。
- （2）增加影响销量的特征变量，用于构建更加完善的回归模型。

八、 小结

通过以上的数据分析，可以得出以下结论：

情况	措施
五台售货机中，位于 E 的售货机销量最好	应重点关注，更加频繁的补进商品
五台售货机中，位于 A 的售货机销量最差	考虑是否更换售货机放置地点，抑或者采取措施增加销量。同时也要及时对未出售的过期的商品进行处理
自动售货机销量最好的时间段之一五六月份，原因应该为天气炎热补充水分	因此应该在此时间段放入更多的饮料类商品
另一个时间段是九月份到十二月份	也应该增加进货量，应对销量增加
一天之中，售货机销量最好的情况是中午和傍晚时分，原因是临近饭点，购买商品作为食物	因此在这些时间段之后应该安排人员进行补货
自动售货机中饮料类商品销量高，较为受欢迎	增补更多的饮料类商品
饮料类商品的毛利润比非饮料类商品的毛利润要多	增补更多的饮料类商品
五台售货机的画像中的茶饮料，碳酸饮料，蛋制品，豆制品，乳制品和肉干类食品销量更多	增补更多的茶饮料，碳酸饮料，蛋制品，豆制品，乳制品和肉干类食品

表 12 情况小结